**Análise de Ativos Financeiros com Regressão**

**Linear e Yahoo Finance**

1. **INTRODUÇÃO**

Este trabalho tem como objetivo aplicar um modelo de regressão linear para prever o preço de fechamento ajustado de ativos financeiros com base em dados históricos obtidos via Yahoo Finance. Foram selecionados três ativos da B3 – **BBDC4**, **LREN3** e **RAIL3** (Bradesco, Lojas Renner e Rumo Logística) – considerando critérios de alta liquidez, disponibilidade de dados (período a partir de 2020, com possibilidade de extensão para 5 anos) e representatividade de setores distintos (financeiro, varejo e logística).

A análise inclui:

* Coleta e pré-processamento dos dados;
* Modelagem;
* Avaliação e discussão.

**Justificativa dos Ativos e Correlação**

**Por que esses ativos foram escolhidos?**

* **Critério de Liquidez:** Cada um dos papéis selecionados (BBDC4, LREN3, RAIL3) apresenta volume médio diário de negociação acima de 1 milhão de ações, garantindo maior liquidez e reduzindo o risco de distorções causadas por baixa negociabilidade.
* **Disponibilidade de Dados Históricos:** Os ativos possuem pelo menos 5 anos de histórico disponível no Yahoo Finance, atendendo ao requisito do trabalho de análise de longo prazo.
* **Setores Diferentes:**
  + **BBDC4 (Banco Bradesco)** – Setor Financeiro
  + **LREN3 (Lojas Renner)** – Setor de Varejo/Consumo
  + **RAIL3 (Rumo)** – Setor de Logística/Infraestrutura

A escolha de setores distintos permite analisar a performance de ativos que reagem de maneira diferente aos ciclos econômicos, diversificando a perspectiva de análise.

A seguir temos os gráficos dos 3 ativos de forma isolada e comparando os mesmos:

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Na comparação podemos notar que todos seguem o mesmo perfil apesar de ser em proporções diferentes: no início da análise estavam em alta no início de 2020 e no início do ano mostram uma queda brusca provavelmente motivados pela Síndrome da COVID que assolou a população mundial afetando vários mercados ao redor do mundo. A seguir mantém-se uma estabilidade de valores baixos durante 2 anos até apresentar uma nova queda sem mostrar uma nova alta. O que nos faz concluir que a queda de mercado é o reflexo de problemas que acontecem ao redor do mundo como: inflação, juros altos, guerras fiscais, instabilidades geradas por guerra, incertezas políticas que causam essa desaceleração de crescimento e várias baixas de mercado.

Interface gráfica do usuário, Gráfico

Descrição gerada automaticamente

A comparação dos 3 slides mostra o comportamento das 3 empresas com uma queda no início da pandemia e sem voltar à alta de antes com porque depois disso, a inflação explodiu (alimentos, energia, transporte). Os bancos centrais começaram a subir os juros (EUA, Brasil, Europa etc.).

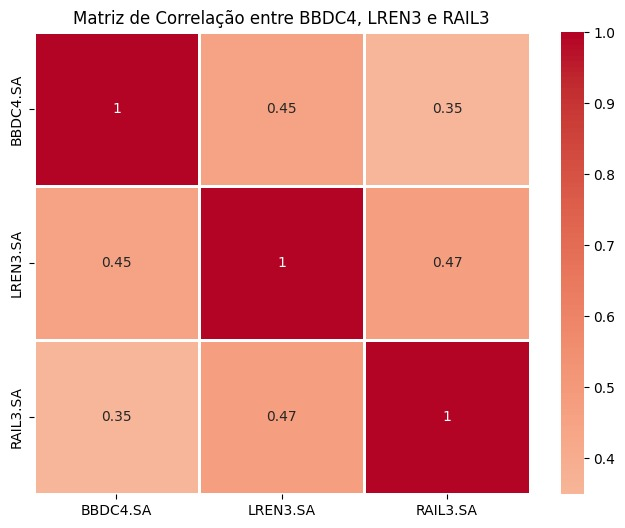
A instabilidade ficou visível no BBDC4 ocasionadas por mudanças de governo (fim do mandato Bolsonaro e entrada do governo Lula) geraram incertezas para o mercado. As desconfianças sobre o controle fiscal, teto de gastos, reforma tributária etc. fizeram com que empresas mais sensíveis a políticas públicas, como bancos sofressem bastante.

Apesar disso, vemos uma queda no consumo, por causa dos juros altos. Um dos fatores é o endividamento das famílias e empresas. E até mesmo empresas do setor varejista perderam valor de mercado como vemos no LENR3.

No Brasil, taxas de juros derrubaram vários segmentos de empresas. O único que mostrou uma alta mais significativa foi o de setor de logística por causa do perfil de compras que fez o mundo e o Brasil (caso da nossa análise) a descobrir as compras pela Internet (RAIL3).

**Qual a correlação histórica entre eles?**

Para avaliar a relação entre os ativos, foi calculada a matriz de correlação dos **retornos diários** no período de 5 anos. A figura a seguir ilustra a **Matriz de Correlação dos Retornos Diários** para BBDC4, LREN3 e RAIL3:



A matriz indica correlações positivas e moderadas entre as ações (0,35 a 0,47). Isso sugere que, apesar de compartilharem influências econômicas, cada ação possui dinâmicas setoriais distintas (BBDC4 no setor bancário, LREN3 no varejo e RAIL3 na logística). Embora não sejam completamente independentes, não são altamente correlacionadas, proporcionando algum nível de diversificação ao serem combinadas em um portfólio, embora não tanto quanto com correlações próximas de zero ou negativas.

**Eles representam setores diferentes?**

Sim. Cada um dos três ativos faz parte de um setor específico: **financeiro**, **varejo** e **logística**. Esse fator atende ao critério de diversificação setorial exigido pelo trabalho, permitindo que a análise capture dinâmicas de mercado mais amplas e reduza a concentração em um único ramo de atividade.

Assim, ao longo do período analisado, cada ativo é afetado de modo distinto por eventos macroeconômicos (por exemplo, alterações na taxa de juros, variações na demanda do varejo, impactos de infraestrutura e exportações), o que se reflete tanto na oscilação de preços quanto nos níveis de correlação.

**2. METODOLOGIA**

**1.1. Dados e Seleção dos Ativos**

* **Ativos:** BBDC4, LREN3 e RAIL3
* **Período:** De 01/01/2020 até a data atual
* **Critérios:**
  + Alta liquidez (volume médio diário elevado)
  + Dados históricos disponíveis
  + Setores diferentes, possibilitando a comparação e diversificação dos comportamentos.

**1.2. Processamento e Visualizações**

Foram realizadas as seguintes etapas:

* **Limpeza dos Dados:** Remoção de dados duplicados e tratamento de dados faltantes.
* **Cálculo dos Retornos:**
  + **Retorno Diário:** Calculado por meio da variação percentual do preço de fechamento.
  + **Retorno Acumulado:** Obtido aplicando o produto acumulado dos retornos diários, iniciando com base 1 em 01/01/2020.

**Resultados Obtidos:**

* **Retorno Total (período):**
  + **BBDC4:** -41,76%
  + **LREN3:** -71,65%
  + **RAIL3:** -34,86%

Observa-se que, embora os três ativos tenham retornos negativos no período analisado, o ativo **RAIL3** apresentou o “melhor” desempenho relativo com uma queda de –34,86%. Esses resultados foram ilustrados por meio de gráficos comparativos dos retornos acumulados ao longo do tempo.

**2. Modelo de Regressão Linear**

Para a previsão dos preços futuros, foi aplicada uma regressão linear utilizando os dados dos 3 ativos. As etapas foram:

* **Definição da Variável Alvo e Feature:**
  + **Variável Alvo (y):** Preço de fechamento (renomeado para “Preço”).
  + **Variável Explicativa (X):** Data convertida para o formato ordinal (representando o tempo).
* **Divisão dos Dados:**
  + **Treino:** Dados anteriores a 180 dias antes do último registro.
  + **Teste:** Últimos 6 meses, para avaliar a performance do modelo.
* **Treinamento:** Aplicação do LinearRegression() do *scikit-learn* para ajustar o modelo com os dados de treino.
* **Avaliação:**
  + Cálculo das métricas de erro: RMSE e MAE para os conjuntos de treino e teste.
  + Estimativa do erro padrão dos resíduos, utilizado para a construção dos intervalos de confiança das previsões futuras.

**2.2. Resultados do Modelo (BBDC4)**

Interface gráfica do usuário, Gráfico, Gráfico de linhas

Descrição gerada automaticamente

O gráfico mostra como as ações de BBDC4 têm se comportado ao longo dos últimos anos, refletindo as oscilações típicas desse mercado. Há uma linha principal (em verde) que serve para indicar uma tendência aproximada, enquanto, no lado direito, aparece uma região em tons de roxo que representa o intervalo de projeção para os próximos meses. A cor vermelha na parte final sugere onde termina o conjunto de dados reais e começa a estimativa do modelo.

* + 1. **Métricas de Avaliação:**
  + **Treino:**
    - RMSE: 1.86
    - MAE: 1.52
    - R²: 0.17
  + **Teste:**
    - RMSE: 1.07
    - MAE: 0.91
    - R²: -0.27
  + **Erro Padrão dos Resíduos:** 1.86
    1. **Previsões Futuras (próximos 6 meses):**

2025-05-07: R$ 12.46 (Intervalo: R$ 8.80 - R$ 16.12)

2025-06-06: R$ 12.41 (Intervalo: R$ 8.74 - R$ 16.07)

2025-07-06: R$ 12.36 (Intervalo: R$ 8.69 - R$ 16.02)

2025-08-05: R$ 12.31 (Intervalo: R$ 8.64 - R$ 15.97)

2025-09-04: R$ 12.26 (Intervalo: R$ 8.59 - R$ 15.92)

2025-10-04: R$ 12.21 (Intervalo: R$ 8.54 - R$ 15.87)

**2.3 - Resultados do Modelo (LREN3)**

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

O gráfico mostra que o preço das ações vem caindo ao longo do tempo, com algumas oscilações de sobe e desce. A linha principal (em verde) indica a trajetória geral que o modelo está apontando, e a região em roxo, no final, é uma estimativa do que pode acontecer nos próximos meses.

Pelos números apresentados, há uma expectativa de que o preço possa ficar na faixa dos 6 reais em maio, descendo para perto de 4 reais por volta de outubro. Porém, o intervalo dessas projeções é bem amplo, chegando a mostrar valores negativos e até acima de 10 reais. Isso quer dizer que há muita incerteza envolvida: o preço pode tanto cair mais quanto ter uma recuperação considerável.

* + 1. **Métricas de Avaliação:**
  + **Treino:**
    - RMSE: 3.46
    - MAE: 2.66
    - R²: 0.83
  + **Teste:**
    - RMSE: 5.01
    - MAE: 4.91
    - R²: -10.51
  + **Erro Padrão dos Resíduos:** 3.51
    1. **Previsões Futuras (próximos 6 meses):**

2025-05-11: R$ 6.54 (Intervalo: R$ -0.36 - R$ 13.44)

2025-06-10: R$ 6.08 (Intervalo: R$ -0.83 - R$ 12.98)

2025-07-10: R$ 5.62 (Intervalo: R$ -1.29 - R$ 12.52)

2025-08-09: R$ 5.16 (Intervalo: R$ -1.75 - R$ 12.06)

2025-09-08: R$ 4.70 (Intervalo: R$ -2.21 - R$ 11.60)

2025-10-08: R$ 4.24 (Intervalo: R$ -2.67 - R$ 11.14)

* 1. **- Resultados do Modelo (RAIL3)**

Interface gráfica do usuário, Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Este gráfico retrata a variação dos preços de RAIL3 ao longo do tempo, junto de uma linha que indica a tendência capturada pelo modelo de regressão linear. Na parte roxa, estão as projeções para os próximos meses, com uma margem de confiança que reflete a incerteza dessas estimativas. Dá para perceber que, embora o preço tenha apresentado altos e baixos, a linha verde sugere uma leve inclinação para cima no período analisado. Isso indica que, se as condições de mercado não mudarem drasticamente, o valor das ações pode continuar por volta dos níveis atuais, com possibilidade de pequenas variações para cima.

**2.4.1. Métricas de Avaliação:**

* + **Treino:**
    - RMSE: 2.29
    - MAE: 1.89
    - R²: 0.03
  + **Teste:**
    - RMSE: 2.59
    - MAE: 2.33
    - R²:-4.46
  + **Erro Padrão dos Resíduos:** 2.30
    1. **Previsões Futuras (próximos 6 meses):**

2025-05-11: R$ 20.86 (Intervalo: R$ 16.35 - R$ 25.37)

2025-06-10: R$ 20.88 (Intervalo: R$ 16.37 - R$ 25.39)

2025-07-10: R$ 20.90 (Intervalo: R$ 16.39 - R$ 25.42)

2025-08-09: R$ 20.93 (Intervalo: R$ 16.41 - R$ 25.44)

2025-09-08: R$ 20.95 (Intervalo: R$ 16.44 - R$ 25.46)

2025-10-08: R$ 20.97 (Intervalo: R$ 16.46 - R$ 25.49)

**3. CONCLUSÕES E RECOMENDAÇÕES**

A escolha dos ativos BBDC4, LREN3 e RAIL3 demonstrou que, mesmo em um cenário de mercado desafiador (retornos negativos no período), os ativos atuam de forma diferenciada – com RAIL3 apresentando o “melhor” desempenho relativo (-34,86%) em comparação com BBDC4 (-41,76%) e LREN3 (-71,65%).

Enquanto a primeira (BBDC4) e a terceira (RAIL3) análises demonstram resultados mais coerentes e intervalos de confiança razoáveis, a segunda (LREN3) análise evidencia limitações significativas da abordagem linear aplicada sem as devidas adaptações. Para trabalhos futuros, é essencial aprofundar as análises dos resíduos, testar alternativas de modelagem e enriquecer o processo com informações adicionais que possam reduzir a incerteza e fornecer previsões mais robustas e realistas.

**Recomendações para Trabalhos Futuros**

* **Inclusão de Novos Features:**

Incluir indicadores técnicos (médias móveis, volatilidade, volume de negociação, RSI, MACD) para melhorar a capacidade preditiva e captar melhor as não-linearidades dos preços.

* **Revisão dos Modelos Utilizados:**

**Validação dos Pressupostos:** Realizar testes rigorosos para verificar homocedasticidade, normalidade e ausência de autocorrelação dos resíduos. Essa etapa é fundamental para garantir que a abordagem linear seja válida.

**Transformações nos Dados:** Considerar transformações (como a logaritmização) para evitar previsões irreais, como valores negativos, principalmente quando se trata de preços de ativos.

* **Exploração de Modelos Alternativos:**

Além da regressão linear, considerar modelos não-lineares (como Random Forest, redes neurais ou modelos ARIMA) que possam capturar melhor os comportamentos dinâmicos do mercado.

* **Validação Rigorosa:**

Realizar validação cruzada ou métodos de *rolling-window* para testar a robustez do modelo em diferentes períodos, principalmente em cenários de alta volatilidade ou crise.

* **Benchmark e Comparação:**

Comparar os resultados com abordagens simples (por exemplo, previsão de que o preço permanecerá estável) para verificar o ganho real na previsão do modelo.

**REFERÊNCIAS E ANEXO**

* **Bibliotecas Utilizadas:**

*yfinance*, *pandas*, *numpy*, *matplotlib*, *seaborn*, *scikit-learn* e *scipy.stats*.

* **Dados:**  
  Dados históricos dos ativos foram obtidos via Yahoo Finance.
* **Anexos:**  
  Incluir gráficos gerados (séries temporais dos preços, retornos acumulados, histogramas dos retornos e preços, curva de regressão com previsões e intervalos de confiança) e o código completo no Jupyter Notebook, conforme exigido.